AT

PATENT ABSTRACTS OF JAPAN

(11)Publication number : **2000-295247**
(43)Date of publication of application : **20.10.2000**

(51)Int.CI. **H04L 12/28**
H04L 12/56

(21)Application number : **11-244911**
(22)Date of filing : **31.08.1999**

(71)Applicant : **NEC CORP**
(72)Inventor : **MARK GOUDORUU**
SUTAFUROSU KORIOPAROSU
SATEISHI RAO

(30)Priority

Priority number	Priority date	Priority country
99 128685	**09.04.1999**	**US**
99 342976	**30.06.1999**	**US**

(54) SCHEDULING METHOD IN QUEUING SYSTEM AND CHUTING BLANKS SWITCH

(57)Abstract:
PROBLEM TO BE SOLVED: To obtain a high stability with a simple operation by discriminating one port having a maximum cell arrival rate and adding a dummy cell to ports except the maximum rate cell arrival port in order to permit the arrival rates with respect to the whole ports to be nearly equal.
SOLUTION: A chuting blanks switch is provided with four input ports 6-11 to 6-14 and four output ports 6-21 to 6-24, for example. The cell arrives at the input ports 6-11 to 6-14 by different rate. A rate discriminator 6.3 discriminates the input/output ports 6-11 to 6-14 and 6-21 to 6-24 of the maximum arrival rate. A filling settler 6.4 settles a dummy flow corresponding to a proper queue to permit the arrival rates with respect to the whole ports 6-11 to 6-14 and 6-21 to 6-24 to be equal. A dummy cell generator 6.5 generates the dummy cell. A scheduler 6.6 executes routing by a heuristic MBM method without considering a queue length.

LEGAL STATUS

[Date of request for examination]	10.07.2000
[Date of sending the examiner's decision of rejection]	
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]	
[Date of final disposal for application]	
[Patent number]	3175735
[Date of registration]	06.04.2001
[Number of appeal against examiner's decision of rejection]	
[Date of requesting appeal against examiner's decision of rejection]	
[Date of extinction of right]	

(19)日本国特許庁（ＪＰ）　(12)**公開特許公報（Ａ）**　(11)特許出願公開番号
特開2000－295247
（P2000－295247A）
(43)公開日　平成12年10月20日(2000.10.20)

(51)Int.Cl.'	識別記号	ＦＩ	ﾃｰﾏｺｰﾄﾞ(参考)
Ｈ０４Ｌ　12/28		Ｈ０４Ｌ　11/20	Ｇ　５Ｋ０３０
12/56			１０２Ｚ　９Ａ００１

審査請求　有　請求項の数18　ＯＬ　（全 16 頁）

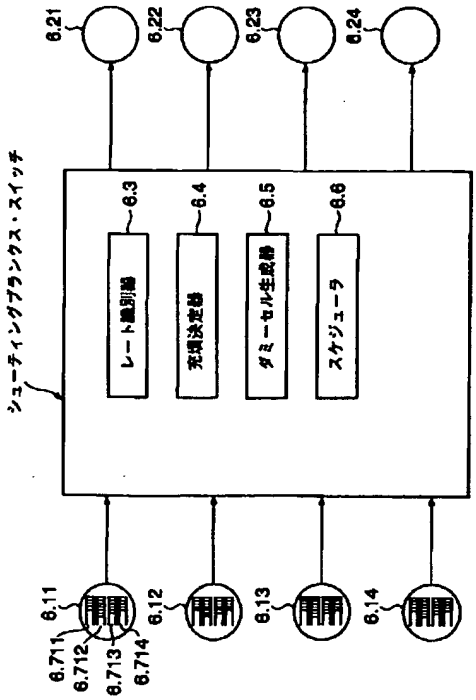
(21)出願番号　特願平11－244911
(22)出願日　平成11年８月31日(1999.8.31)
(31)優先権主張番号　０９／３４２９７６
(32)優先日　平成11年６月30日(1999.6.30)
(33)優先権主張国　米国（US）
(31)優先権主張番号　６０／１２８６８５
(32)優先日　平成11年４月９日(1999.4.9)
(33)優先権主張国　米国（US）

(71)出願人　000004237
日本電気株式会社
東京都港区芝五丁目７番１号
(72)発明者　マーク・ゴウドルー
アメリカ合衆国、ニュージャージー
08540　プリンストン、４　インディペンデンス　ウエイ、エヌ・イー・シー・ユー・エス・エー・インク内
(74)代理人　100097157
弁理士　桂木　雄二

最終頁に続く

(54)【発明の名称】　キューイングシステムにおけるスケジューリング方法及びシューティングプランクス・スイッチ

(57)【要約】
【課題】　ＭＷＢＭ（最大重み２部マッチング）スケジューラの優れた安定性を達成するとともに、より単純なＭＢＭ（最大サイズ２部マッチング）またはＭＢＭベースのヒューリスティックスを実行するスイッチのためのスケジューリング方法を提供する
【解決手段】　スイッチのＭ個の入力ポートとＮ個の出力ポートの間でセルがルーティングされる。セルは、ポートのサブセットごとに異なるレートで到着する。Ｍ個の入力ポートはそれぞれＮ個のキューを有することにより、全部でＭ×Ｎ個のキューが存在する。スケジューリング方法は、まず、１つのポートを、最大のセル到着レートを有するとして識別する。次に、すべてのポートに対する到着レートがほぼ等しくなるように、最大のセル到着レートを有するとして識別されたポート以外のすべてのポートにダミーセルを追加する。



【特許請求の範囲】

【請求項１】　Ｎ個の入力キューを１つのグループとしてＭ個のグループを有し、前記Ｎ個の入力キューはそれぞれ、Ｎ個のサービスのうちの相異なる１つを要求し、Ｍ及びＮは自然数であり、従って前記システムはＭ×Ｎ個のキューを有するシステムのためのスケジューリング方法において、

（ａ）Ｍ個のグループ及びＮ個のサービスに対するエントリの到着レートを判定するステップと、

（ｂ）前記Ｍ個のグループ及びＮ個のサービスのうち最大到着レートを有する１つを識別するステップと、

（ｃ）前記Ｍ個のグループ及びＮ個のサービスのうち、前記ステップ（ｂ）で識別されたグループまたはサービス以外のすべてのグループ及びサービスに対して、前記Ｍ個のグループ及びＮ個のサービスに対するエントリの到着レートがほぼ等しくなるように、ダミーエントリを追加するステップと、

（ｄ）入力キューの長さを考慮しないスケジューリング法を使用するステップと、

からなることを特徴とするスケジューリング方法。

【請求項２】　Ｍ個の入力ポート及びＮ個の出力ポートを有し、前記Ｍ個の入力ポートと前記Ｎ個の出力ポートの間でセルがルーティングされ、前記セルはポートのサブセットごとに異なるレートで到着し、前記Ｍ個の入力ポートはそれぞれＮ個のキューを有することにより全部でＭ×Ｎ個のキューが存在するスイッチのルーティング方法において、

（ａ）最大のセル到着レートを有する１つのポートを識別するステップと、

（ｂ）すべての入力ポート及び出力ポートに対する到着レートがほぼ等しくなるように、最大のセル到着レートを有するとして識別された入力ポートまたは出力ポート以外のすべての入力ポート及び出力ポートにダミーセルを追加するステップと、

（ｃ）キュー長を考慮しないスケジューリング法を使用するステップとからなり、

前記ステップ（ａ）において、前記Ｍ個の入力ポートのうちの１つの入力ポートに対する到着レートは、当該１つの入力ポートに関連するＮ個のキューの到着レートの和であり、前記Ｎ個の出力ポートのうちの１つの出力ポートに対する到着レートは、前記Ｍ個の入力ポートの各入力ポートから当該１つの出力ポートへの到着レートの和である、

ことを特徴とするルーティング方法。

【請求項３】　前記ステップ（ｂ）は、

（ｂ．１）すべてのポートに対する到着レートが等しくなるように適当なキューに対するダミーフローを決定して充填を実行するステップと、

（ｂ．２）前記ステップ（ｂ．１）で決定された到着レートでダミーセルの生成を実行するステップと、

からなることを特徴とする請求項２記載の方法。

【請求項４】　前記ステップ（ｂ．１）の充填は、

（ｂ．１．１）ポートのうちで最大到着レートΛのものを識別するステップと、

（ｂ．１．２）入力ポートｉと出力ポートｊの間のパケット伝送レートγ_{ij}を、初期伝送レートα_{ij}に初期化するステップと、

（ｂ．１．３）入力ポートポインタ及び出力ポートポインタを１に初期化するステップと、

（ｂ．１．４）入力ポートｉの到着レートR_i及び出力ポートｊの到着レートS_jがΛより小さくなるまで、前記入力ポートポインタ及び前記出力ポートポインタをインクリメントするステップと、

（ｂ．１．５）ｉとｊの間のフローに対するダミーセルを、R_iがS_j以上である場合はレート$\beta_{ij}=\Lambda-R_i$で、それ以外の場合はレート$\beta_{ij}=\Lambda-S_j$で、追加するステップと、

（ｂ．１．６）目標到着レートγ_{ij}を$\gamma_{ij}+\beta_{ij}$に更新するステップと、

（ｂ．１．７）すべてのポートについて前記ステップ（ｂ．１．４）～（ｂ．１．６）を繰り返すステップと、

からなることを特徴とする請求項３記載の方法。

【請求項５】　前記ステップ（ｂ．２）の生成は、目標到着レートが満たされるように各入力ポートにおいてダミーセルを生成し且つ実セルが入力ポートに到着しているときには当該実セルを対応するダミーセルにピギーバックすることによって実行され、

各キューは、該キューにおいて伝送のためにキューイングしているダミーセルの数に対応するクレジット値を保持し、

スケジューラは、前記クレジット値が１以上であるときに、特定のキュー内のセルについて知らされ、当該特定のキューが選択されると前記クレジット値は１だけデクリメントされることを特徴とする請求項３記載の方法。

【請求項６】　ダミーセルは実際に生成されるのではなく、各入力キューが、前記クレジット値に等しい数であってｉからｊへの伝送のためにキューイングしているダミーセルの数に対応する数d_{ij}を保持し、

前記数d_{ij}及び前記クレジット値は、各時間ステップごとに、確率γ_{ij}でインクリメントされることを特徴とする請求項５記載の方法。

【請求項７】　ｉとｊの間のコネクションが許可されたときに、実セルがｉとｊの間の伝送のためにキューイングされている場合には、該実セルが伝送され、d_{ij}が１だけデクリメントされることを特徴とする請求項６記載の方法。

【請求項８】　ダミーセルは生成されるのではなく、各入力キューが数c_{ij}を保持し、

前記数c_{ij}は、前記クレジット値に等しく、

前記数c_{ij}及び対応するクレジット値は、ｉからｊへの伝送のためにキューイングされるダミーセルの数に対応する浮動小数点数であり、
前記数c_{ij}及び前記クレジット値は、各時間ステップごとに、０と１の間の小数をとるγ_{ij}だけインクリメントされることを特徴とする請求項５記載の方法。
【請求項９】　ｉとｊの間のコネクションが許可されたときに、実セルがｉとｊの間の伝送のためにキューイングされている場合には、該実セルが伝送されることを特徴とする請求項８記載の方法。
【請求項１０】　$c_{ij} \geqq 1$の場合、または、実セルがｉとｊの間のキューに存在する場合、該キューは、キューｉとｊの間でセルを伝送することができることをスケジューラに通知することを特徴とする請求項８記載の方法。
【請求項１１】　シューティングブランクス・スイッチにおいて、
Ｍ＋Ｎ個のポートと、レート識別器と、充填決定器と、ダミーセル生成器と、スケジューラと、を有し、
前記ポートは、Ｍ個の入力ポート及びＮ個の出力ポートからなり、該Ｍ個の入力ポートと該Ｎ個の出力ポートの間でセルがルーティングされ、該セルは、ポートのサブセットごとに異なるレートで到着し、前記Ｍ個の入力ポートはそれぞれＮ個のキューを有することにより、全部でＭ×Ｎ個のキューが存在し、
前記レート識別器は、最大のセル到着レートを有する１つのポートを識別し、前記Ｍ個の入力ポートのうちの１つの入力ポートに対する到着レートは、該１つの入力ポートのＮ個のキューに対する到着レートの和であり、前記Ｎ個の出力ポートのうちの１つの出力ポートに対する到着レートは、前記Ｍ個の入力ポートのそれぞれから該１つの出力ポートへの到着レートの和であり、
前記充填決定器は、すべてのポートに対する到着レートが等しくなるように適当なキューに対するダミーフローを決定し、
前記ダミーセル生成器は、前記充填決定器によって決定された到着レートでダミーセルを生成し、
前記スケジューラはキュー長を考慮しない、
ことを特徴とするシューティングブランクス・スイッチ。
【請求項１２】　前記生成器は、目標到着レートが満たされるように各入力ポートにおいてダミーセルを生成し、実セルが入力ポートに到着しているときには該実セルを対応するダミーセルにピギーバックし、
各キューは、該キューにおいて伝送のためにキューイングされているダミーセルの数に対応するクレジット値を保持し、
前記スケジューラは、前記クレジット値が１以上であるときに、特定のキュー内のセルについて知らされ、該特定のキューが選択されると前記クレジット値は１だけデクリメントされる、
ことを特徴とする請求項１１記載のシューティングブランクス・スイッチ。
【請求項１３】　ダミーセルは実際に生成されるのではなく、各入力キューが数d_{ij}を保持し、前記数d_{ij}は前記クレジット値に等しく、ｉからｊへの伝送のためにキューイングされるダミーセルの数に対応し、
前記数d_{ij}は、各時間ステップごとに、目標レートγ_{ij}に等しい確率でインクリメントされる、
ことを特徴とする請求項１１記載のシューティングブランクス・スイッチ。
【請求項１４】　ｉとｊの間のコネクションが許可されたときに、実セルがｉとｊの間の伝送のためにキューイングされている場合には、該実セルが伝送されることを特徴とする請求項１３記載のシューティングブランクス・スイッチ。
【請求項１５】　ダミーセルは生成されるのではなく、各入力キューは数c_{ij}を保持し、
前記数c_{ij}は、前記クレジット値に等しく、
前記数c_{ij}及び対応するクレジット値は、ｉからｊへの伝送のためにキューイングされるダミーセルの数に対応する浮動小数点数であり、
前記数c_{ij}は、各時間ステップごとに、０と１の間の小数をとるγ_{ij}だけインクリメントされる、
ことを特徴とする請求項１２記載のシューティングブランクス・スイッチ。
【請求項１６】　ｉとｊの間のコネクションが許可されたときに、実セルがｉとｊの間の伝送のためにキューイングされている場合には、該実セルが伝送されることを特徴とする請求項１５記載のシューティングブランクス・スイッチ。
【請求項１７】　$c_{ij} \geqq 1$の場合、または、実セルがｉとｊの間のキューに存在する場合、前記キューはｉとｊの間でセルを伝送することができることを前記スケジューラに通知し、c_{ij}は１だけデクリメントされることを特徴とする請求項１６記載のシューティングブランクス・スイッチ。
【請求項１８】　スイッチのサブセットが請求項１１記載のシューティングブランクス・スイッチであるようなスイッチを有することを特徴とするネットワーキングシステム。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】本発明は、複数のサービスを要求する複数のキューを有するキューイングシステムのスケジューリング方法に係り、特に入力キュースイッチにおける高いスループットを達成するためのシューティングブランクス(shooting blanks)方法に関する。
【０００２】
【従来の技術】いくつかのアプリケーションは、複数の

サービスを要求する複数のキューの割当てを必要とする。本発明は、そのようなキューイングシステムにおけるスケジューリング技術の改良に関する。

【０００３】［グラフ理論の用語］本発明の理解を容易にするために、ここでいくつかのグラフ理論用語を定義する。

【０００４】2部グラフとは、図１に例示するように、複数の頂点を１．１～１．４と１．５～１．８という２つの集合に分割することができて、辺１．９がこれらの２つの集合の間にのみ位置するような無向グラフをいう。サービスの集合を要求するキューの集合は、このような２部グラフによって表現することができる。辺は、第１の頂点と第２の頂点の間のキューを表す。２頂点間の辺の存在は、対応するキューが第２の頂点によって表されるサービスを要求することを示す。

【０００５】２部マッチングとは、どの頂点も複数の辺の端点にならないような辺の部分集合である。マッチングを求めることは、キューと、要求されるサービスとの間の置換（あるいは並び替え）の１つを選択することに対応する。

【０００６】・最大（サイズ）マッチングとは、可能な最大数の辺からなるマッチングである。ｎ及びｍを、それぞれ、現在の状態を表す基礎となるグラフの頂点及び辺の数とすると、Ｎ個のサービスを要求するN^2個のキューを表すＮ×Ｎシステムの場合、ｎ＝２Ｎであり、ｍ≦N^2である。

【０００７】・最大重みマッチングとは、グラフ内の辺に関連づけられた重みも考慮に入れたマッチングをいう。本発明では、重みは、対応するキューのサイズに等しくする。

【０００８】［システムモデル］背景的知識及び本発明をよりよく理解するため、次に、システムモデルについて説明する。当業者には明らかなように、システムモデルは、ここで説明する事項の理解を簡単にするためにのみ用いられるものであって、本発明の技術的範囲を制限するように解釈してはならない。

【０００９】ここでは、Ｎ個の入力ポート及びＮ個の出力ポートを有するクロスバスイッチを含む交換ファブリックを使用するが、本発明は、Ｍ個の入力ポート及びＮ個の出力ポートを有するスイッチにも適用可能である。このクロスバスイッチは、１つの時間ステップ（以下、１ステップという。）で、固定サイズのパケットの任意の置換（並び替え）をルーティングすることができる。１ステップで、入力ポートを出ること、または、出力に到達することが可能なセルは１個だけである。従って、多数個のセルは、入力で待合せ（キューイング）されることがある。入力ポートごとに１個のキューを使用すると、ＨＯＬ(head-of-line)ブロッキング効果を引き起こし、スループットをひどく制限する可能性がある。このようなブロッキング効果について詳細な情報は、Mark J. Karol, Michael G. Hluchyj, and Samuel P. Morgan, "Input versus output queuing on a space-division packet switch", IEEE Transactions on Communications, COM-35(12):1347-1356,December 1987、に記載されている。

【００１０】さまざまな研究グループが、入力ポートごとにＮ個のキュー（Ｎ個の出力ポートのそれぞれに対して１個ずつのキュー）を使用することを提唱している。例えば、

・Thomas E. Anderson, Susan S. Owicki, James B. Saxe, and Charles P. Thacker, "High-speed switch scheduling for local-area networks", ACM Transactions on Computer Systems, 11(4):319-352, November 1993

・M. Karol, K. Eng, and H. Obara, "Improving the performance of input-queued ATM packet-switching", In Proceedings IEEE INFOCOM' 92, p.110-115, 1992

・Nicholas William McKeown, "Scheduling Algorithms for Input-Queued Cell Switches", PhD thesis, University of California at Berkeley, 1995

を参照のこと。

【００１１】これらに記載されている技術（仮想出力キューイング（ＶＯＱ：virtual output queuing）という。）は、背景的知識として、本発明の好ましい実施例を説明するために用いられるモデルである。直観的には、ＶＯＱによれば、スイッチは、置換を選択する際に優れたフレキシビリティを有する。図２は、Ｎ＝４のＶＯＱスイッチの一例を示す模式的構成図である。

【００１２】入力ポートｉから出力ポートｊに行く必要があるパケットの到着レートをα_{ij}で表す。レート５０％とは、平均として、１ステップおきに１個のパケットが、入力ポートｉから出力ポートｊに送られることを意味する。パケットが入力ポートｉに到着するレートをR_iとする。これは$\Sigma_j \alpha_{ij}$に等しい。パケットが出力ポートｊに到着するレートをS_jとする。これは$\Sigma_i \alpha_{ij}$に等しい。なお、この総和において、ｉはすべての入力ポートを表し、ｊはすべての出力ポートを表す。正式な用語ではないが、際限なく成長する入力ポートがない場合にスイッチは安定であるという。すべてのｉ、ｊについて、R_i, S_j＜１００％のとき、トラフィックは許容されるという。

【００１３】［ルータスイッチとスケジューリング］このようなキューイングシステムの重要な例は、インターネットのようなアプリケーションで用いられるルータスイッチである。本発明は、２部グラフを用いて表現可能なキューイングシステムの重要な例としてスイッチを使用する。ただし、本発明は、スイッチに限定されるものではなく、２部グラフを用いて表現可能な任意のキューイングシステムに適用可能であることに注意すべきである。

【００１４】インターネットの発展は、ルータ及びスイ

ッチの需要をますます増大させ続けている。毎秒数十ないし数百ギガビットの広帯域を処理可能で、それとともに多数の入出力ポートを有するスイッチ設計法が現在は利用可能である。ネットワークシステムにおいて、セル（あるいはパケット）は、いくつかの入力ポートを通じてスイッチに到着する。これらのセルは、スイッチによって使用されるスケジューリング方法に基づいて、1つまたはいくつかの出力ポートを通じてルーティングされる。なお、本明細書では、パケットとセルという用語は交換可能であり、本発明は、いかなる種類のセルあるいはパケットにも制限されるものではない。

【００１５】広帯域スイッチの設計者は、交換ファブリックを通じてセルをスケジューリングする有効な技術を提供するという問題に直面している。現在、理論的な観点から、知られている最良のアプローチは、従来の最大重み2部マッチング（ＭＷＢＭ：Maximum Weight Bipartite Matching）を利用するものである。入出力ポートが１００％容量で動作する必要がない限り、かなり一般的な確率的仮定のもとで、ＭＷＢＭは安定であることが、Tassiulas and Ephremidesと、McKeown, Ananthara m, and Walrandとによって独立に示された。（入力ポートまたは出力ポートに関連する）いずれのキューも無制限に成長しないとき、スケジューリング方法は安定であるという。ＭＷＢＭに関するその他の情報については、

・Leandros Tassiulas and Anthony Ephremides, "Stability properties ofconstrained queuing systems and scheduling policies for maximum throughput in multihop radio networks", IEEE Transactions on Automatic Control,37(12):1936-1948, December 1992

・Nick McKeown, Venkat Anantharam, and Jean Walrand, "Achieving 100% throughput in an input-queued switch", In Proceedings IEEE INFOCOM' 96, p.296-302, San Francisco, CA, March 1996

に記載されている。入出力ポートの負荷がフルでない場合のトラフィックパターンは許容されるとみなされる。

【００１６】しかしながら、ＭＷＢＭの既知の最良の方法の実行時間は、Ｏ（Ｎ^{2}ｌｏｇＮ＋Ｎｍ）及びＯ（Ｎ$^{1/2}$ｍｌｏｇＮＣ）である。ただし、Ｎはスイッチ内の入力ポートの数であり、ｍ≦Ｎ^{2}は空でないキューの数であり、Ｃは最大キューサイズである。ＭＷＢＭ法の実行時間に関する詳細な情報については、

・M. L. Fiedman and R. E. Tarjan, "Fibonacci heaps and their uses in improved network optimization algorithms", Journal of the ACM, 34:596-615, 1987

・H. N. Gabow and R. E. Tarjan, "Faster scaling algorithms for networkproblems", SIAM Journal on Computing, 18:1013-1036, 1989

に記載されている。さらに、これらの方法は本質的にシーケンシャルである。置換ごとに伴う実行時間のために、ＭＷＢＭの完全な実現は非実際的である。ＭＷＢＭの完全な実現は、セル時間あたり１つの置換計算を必要とする広帯域スイッチでは可能性はあまりないと考えられている。置換はマッチングと等価である。

【００１７】これに対して、最大（サイズ）2部マッチング（ＭＢＭ：Maximum (Size) Bipartite Matching）に基づくスケジューリングアプローチには次の2つの主要な利点がある。

【００１８】（１）第1に、最良の既知のＭＢＭ法の実行時間はＯ（Ｎ$^{1/2}$ｍ）である。ＭＢＭの実行時間に関するさらに詳細な情報については、

・J. E. Hopcroft and R. M. Karp, "An $n^{5/2}$ algorithm for maximum matching in bipartite graphs", SIAM Journal on Computing, 2:225-231, 1973

・E. A. Dinitz, "Algorithm for solution of a problem of maximum flow in networks with power estimation", Soviet Math. Dokl., 11:1277-1280, 1970

・S. Even and R. E. Tarjan, "Network flow and testing graph connectivity", SIAM Journal on Computing, 4:507-518, 1975

に記載されている。

【００１９】（２）第2に、この方法の入力は、各キューが空であるか否かについての知識（Ｎ^{2}ビット）しか必要としない。これに対して、ＭＷＢＭ法は、各キューの長さについての知識（Ｎ^{2}個の整数。ただし、整数のサイズには上限がない。）を必要とする。

【００２０】しかしながら、ＭＷＢＭと比べた場合のＭＢＭの欠点は、ＭＢＭは、すべての許容トラフィックパターンに対して安定であるとは限らないことである。特に、McKeown et al.は、許容されるがＭＢＭのもとで安定ではないＮ＝3の単純なトラフィックパターンを示している（Nick McKeown, Venkat Anantharam, and Jean Walrand, "Achieving 100% throughput in an input-queued switch", In Proceedings IEEE INFOCOM' 96, p.296-302, San Francisco, CA、を参照）。

【００２１】しかし、ＭＢＭが安定になるような許容トラフィックパターンの興味深いクラスがあることが予想される。このクラスのトラフィックは、平衡(balanced)トラフィックパターンである。平衡トラフィックパターンの場合、すべての入出力ポートの負荷は均一である。以下の事実は、上記の予想を正当化する。

【００２２】１．　ＭＢＭは、平衡スケジューリング問題のオフラインバージョンに対して１００％の利用率を提供することが知られている。各入力ポートがちょうどＴ個のセルを送出することを必要とし、各出力ポートがちょうどＴ個のセルを受け取ることを必要としている場合、ＮＴ個のすべてのセルをちょうどＴ回の置換でルーティングすることが可能である。以下を参照のこと：

・G. Birkhoff, "Tres observaciones sobre el algebra lineal", Univ. Nac. Tuceman Rev. Ser. A5, p.147-150, 1946;

・Shizhao Li and Nirwan Ansari, "Input-queued switching with QoS guarantees", In Proceedings IEEE INFOCOM' 99, volume 3, p.1152-1159, March 1999)。従って、MBM法をT回使用して置換を選択することができる。

【0023】2. 現在、MBMアプローチの既知のすべてのシミュレーションは、トラフィックフローが平衡であるときは必ず100%負荷まで安定である。シミュレーションは、不平衡フローの場合にのみ、不安定性を示している。

【0024】置換選択問題は、グラフ理論用語でモデル化することができる。2部グラフとは、図1に示すように、頂点を2つの集合に分割することが可能であり、各辺がこれらの2つの集合間にあるような無向グラフである。2部グラフは、スイッチを表現するために用いることができる。この場合、各入力ポートは一方の集合のノード(節点、すなわち頂点)によって表され、各出力ポートは、他方の集合のノードによって表される。2頂点間に辺が存在することは、スイッチ内の対応する入力ポートと出力ポートとの間でルーティングされるべきセルが存在することを示す。極大マッチングとは、マッチングを維持したままで辺を追加することができないようなマッチングのことである。

【0025】文献に現れているほとんどのスケジューリング方法は、2つのカテゴリに分けられる。キューのサイズを考慮するもの(MWBMによって例示される)と、考慮しないもの(MBMによって例示される)である。

【0026】MWBMアプローチでは、2部グラフの辺に重み値が割り当てられる。一般に、入力ポートiから出力ポートjへの辺には、入力ポートiにキューイングされた、出力ポートjを宛先とするパケットの数q_{ij}に等しい重みが割り当てられる。その後、得られたグラフにおいて最大重みマッチングを計算する。前述のように、MWBMアプローチは、トラフィックパターンが不平衡のときでも安定性を備える。しかし、既知の最良の方法の実行時間は、O($N^2 \log N + Nm$)及びO($N^{1/2} m \log NC$)である。さらに、これらの方法は本質的にシーケンシャル(直列的、逐次的)であり、良好なハードウェア実現が困難である。MWBMを近似しようとするヒューリスティックスも提案されているが、それらもまた、シーケンシャル計算のボトルネックによって制限されるようである。以下参照のこと:

・Adisak Mekkittikul and Nick McKeown, "A practical scheduling methodto achieve 100% throughput in input-queued switches", In Proceedings IEEE INFOCOM' 98, p.792-799, San Francisco, CA, March 1998;

・Anthony C. Kam and Kai-Yeung Siu, "Linear complexity methods for bandwidth reservations and delay guarantees in input-queued switches with no speedup", IEEE International Conference on Network Protocols, Austin,Texas, October 1998)。

【0027】MBMアプローチでは、可能な最大数のパケットがタイムスロットごとにルーティングされる。MWBM法よりも漸近的に優れてはいるが(既知の最良のMBM法の実行時間はO($N^{1/2} m$)でしかない)、MBMは依然として、広帯域スイッチの速度制約のもとでは実現が困難な方法である。PIMやSLIPのようないくつかのヒューリスティック近似が開発されている。これらは、より簡単な実現が可能であり、最大サイズマッチングに近似しようとするものである。以下参照のこと:

・Nick McKeown, Martin Izzard, Adisak Mekkittikul, William Ellersick,and Mark Horowitz, "The Tiny Tera: A packet core switch", In Hot Interconnects IV, p.161-173, Stanford University, August 1996;

・Nicholas William McKeown, "Scheduling Algorithms for Input-Queued Cell Switches", PhD thesis, University of California at Berkeley, 1995;

・Thomas E. Anderson, Susan S. Owicki, James B. Saxe, and Charles P. Thacker, "High-speed switch scheduling for local-area networks", ACM Transactions on Computer Systems, 11(4):319-352, November 1993)。これらのヒューリスティックアプローチもまた、不平衡の許容トラフィックパターンに対しては不安定性を示すことがある。

【0028】[MBMに対するスケジューリングヒューリスティックス]当業者には理解されるように、MBM法についていくつかのヒューリスティックスが使用可能である。ここで説明するMBM関連のヒューリスティックスは、後述するセクションIVDに記載する実験で、さまざまなMBM法の包括的な比較研究のために用いられる。

【0029】PIM: 並列反復マッチング(Parallel Iterative Matching)は、平均としてO($\log n$)で極大マッチングに収束するヒューリスティックスである(Thomas E. Anderson, Susan S. Owicki, James B. Saxe, and Charles P. Thacker, "High-speed switch scheduling for local-area networks", ACM Transactions on Computer Systems, 11(4):319-352, November 1993、を参照)。各イタレーションごとに、マッチングしていない入力が、キューイングされているセルを有するあらゆる出力に対して提案(要求)をする。マッチングしていない出力は、要求を受け取った場合、すべての要求にわたって一様ランダムに選択を行うことによって1つの要求を受け入れる。1つの入力が複数の許可を受け取った場合、その入力は、それらのうちからランダムに1つを選択する。

【0030】RND: RNDヒューリスティックスはPIMに類似している。相違点は、マッチングしていな

いあらゆる入力iは、iがキューイングされているセルを有するようなランダムに選択された出力ポートjに対する1つの要求のみを行う点である。

【0031】RR-RND： ラウンドロビン法RR-RNDでは、与えられたタイムスロットtに対するスケジューリングは、固定された優先スケジュールによって決定される。最高の優先度の入力i_1は、それがキューイングされたセルを有するような出力のうちからランダムに選択を行う。次に高い優先度の入力i_2は、取られていない出力のうちから次の選択を行い、以下同様である。次のタイムスロットt+1において、入力ポート優先度は、Nを法として1だけインクリメントされる。この優先方式の利点は、パイプライン機構を有するハードウェアで容易に実現可能なことである。

【0032】SLIP： SLIPは、McKeownによって開発され、広範に解析されている（Nicholas William McKeown, "Scheduling Algorithms for Input-Queued CellSwitches", PhD thesis, University of California at Berkeley, 1995、を参照）。各入力ポート及び出力ポートは、別々の優先度ホイール（回転盤）を有する。1つの入力は、それがキューイングされたセルを有するようなあらゆる出力に1つずつ要求を送る。1つの出力jは、自己の優先度ホイールに基づいて、ポートiからの要求を受け入れることを選択する。iによって受け入れられた場合、jにおける優先度ホイールは、（i+1）modNにセットされる。入力ポートは、優先度ホイールに基づいて許可を受け入れ、自己のホイールを同様に回す。

【0033】SHAKEUP： SHAKEUP法は、米国特許出願第09/342，975号（本特許出願と同日出願の特願平11-244648号）において本発明の発明者によって導入されたものである。SHAKEUPは、それ自体でも、他の方法と組み合わせても使用可能であり、大きなパフォーマンス利得を達成することが示されている（前記特願平11-244648号を参照のこと）。この方法は、2部グラフで、何らかの初期マッチングを仮定する。初期マッチングは、例えば、他のヒューリスティックスにより生成される。基本的な考え方は、A内のマッチングしていない各頂点は、たとえ既存のマッチングを除去することになるとしても、その頂点自身がマッチングを強制されるということである。具体的には、少なくとも1本の辺の端点である、集合A内のマッチングしていない各頂点xは、端点となっている辺のうちの1つ（例えば、（x，y））を一様ランダムに選択する。y∈Bがマッチングしていない場合、（x，y）をマッチングに追加する。yが既に辺（z，y），z∈Aによりマッチングしている場合、（z，y）をマッチングから除去し、（x，y）を追加する。最後に、SHAKEUP中に、A内の複数の頂点がB内の同じ頂点とマッチングしようとする場合、A内の競合する頂点のうちの1つをランダムに勝者として選択する。ここで、SHAKEUPは、並列に実現されておらず、A内のマッチングしていない頂点についてランダムに選択された順序で実行されると仮定する。SHAKEUPは、少なくとも初期マッチングと同じ大きさのマッチングを返す（SHAKEUPに関する詳細な実験及び理論的正当化については、前記米国特許出願及び特願平11-244648号を参照）。

【0034】なお、PIM、RND及びSLIPのヒューリスティックスは、反復ヒューリスティックスのクラスに入る。反復ヒューリスティックスは、1回の実行（イタレーション）で極大マッチングを見つけることは保証されない。従って、複数回反復することにより、置換のサイズが増大する可能性がある。

【0035】

【発明が解決しようとする課題】前述のように、MBM及びMWBMのアプローチはいずれもそれぞれの欠点を有する。従って、MWBMの安定性を有するとともに、MBMのような低い実行時間要求条件を有するスケジューリング方法が、特に広帯域スイッチを実現するために所望される。

【0036】本発明の目的は、一般的なトラフィックパターンに対して高いパフォーマンスのスケジューリング方法を提供することにある。

【0037】特に、本発明の目的は、MWBMスケジューラの優れた安定性を達成すると共に、より単純なMBMまたはMBMベースのヒューリスティックスを実行する、スイッチのスケジューリング方法を提供することにある。

【0038】

【課題を解決するための手段】本発明のアプローチは、シューティングブランクスと呼ばれる。その主な考え方について以下で説明する。シューティングブランクス法では、すべての入出力が同じ到着レートΛになるように、ダミーパケットを生成することによって軽い負荷のポートを通るトラフィックを増大させる。ダミーパケット生成をした後、MBMまたはMBM関連のヒューリスティックスを適用して、ルーティングされるべき置換を計算する。軽負荷のポートはダミートラフィックに対応する辺の端点に相当するため、この方法は、いわば騙された形で、これらをマッチングに含めるように選択する。

【0039】本発明は、コンピュータ実装されたスケジューリング方法として、ルーティングに用いられる方法として、シューティングブランクス・スイッチとして、またシューティングブランクス・スイッチを用いたネットワークシステムとして、実現される。

【0040】本発明によれば、N個の入力キューを1つのグループとしてM個のグループを有するシステムのための、コンピュータにより実現されたスケジューリング

方法が提供される。前記N個の入力キューはそれぞれ、N個のサービスのうちの相異なる1つを要求する。M及びNは自然数である。従って、システムは、M×N個のキューを有することになる。前記スケジューリング方法は、M個のグループ及びN個のサービスに対するエントリの到着レートを判定するステップと、前記M個のグループ及びN個のサービスのうち最大到着レートを有する1つを識別するステップと、前記M個のグループ及びN個のサービスのうち、識別されたグループまたはサービス以外のすべてのグループ及びサービスに、前記M個のグループ及びN個のサービスに対するエントリの到着レートがほぼ等しくなるように、ダミーエントリを追加するステップと、入力キューの長さを考慮しないスケジューリング法を使用するステップとからなる。

【0041】本発明の他の側面によれば、ポートを有するスイッチにおけるルーティング方法が提供される。前記ポートは、M個の入力ポート及びN個の出力ポートのうちの1つであり、該M個の入力ポートと該N個の出力ポートの間でセルがルーティングされる。前記セルは、ポートのサブセットごとに異なるレートで到着する。前記M個の入力ポートはそれぞれN個のキューを有することにより、全部でM×N個のキューが存在する。前記ルーティング方法は、1つのポートを、最大のセル到着レートを有するとして識別するステップと、すべての入力ポート及び出力ポートに対する到着レートがほぼ等しくなるように、最大のセル到着レートを有するとして識別された入力ポートまたは出力ポート以外のすべての入力ポート及び出力ポートにダミーセルを追加するステップと、キュー長を考慮しないスケジューリング法を使用するステップとからなり、前記M個の入力ポートのうちの1つの入力ポートに対する到着レートは、該1つの入力ポートのN個のキューに対する到着レートの和であり、前記N個の出力ポートのうちの1つの出力ポートに対する到着レートは、前記M個の入力ポートのそれぞれから該1つの出力ポートへの到着レートの和である。

【0042】好ましくは、ダミーセルは、すべてのポートに対する到着レートが等しくなるように適当なキューに対するダミーフローを決定して充填するステップと、決定された到着レートでダミーセルを生成するステップと、からなるプロセスを使用して追加される。

【0043】好ましくは、充填は、ポートのうちで最大到着レートΛのものを識別するステップと、入力ポートiと出力ポートjの間のパケット伝送レートγ_{ij}を、初期伝送レートα_{ij}に初期化するステップと、入力ポートポインタ及び出力ポートポインタを1に初期化するステップと、入力ポートiの到着レートR_i及び出力ポートjの到着レートS_jがΛより小さくなるまで、前記入力ポートポインタ及び前記出力ポートポインタをインクリメントするステップと、iとjの間のフローに対するダミーセルを、R_iがS_j以上である場合はレート$\beta_{ij}=\Lambda-R_i$で、それ以外の場合は$\beta_{ij}=\Lambda-S_j$で、追加するステップと、目標到着レートγ_{ij}を$\gamma_{ij}+\beta_{ij}$に更新するステップと、すべてのポートについて上記ステップを繰り返すステップと、からなるプロセスを使用して実行される。

【0044】好ましくは、生成は、目標到着レートが満たされるように各入力ポートにおいてダミーセルを生成するステップと、実セルが入力ポートに到着しているときには該実セルを対応するダミーセルにピギーバックするステップと、によって実行され、各キューは、該キューにおいて伝送のためにキューイングされているダミーセルの数に対応するクレジット値を保持し、スケジューラは、前記クレジット値が1以上であるときに、特定のキュー内のセルについて知らされ、該特定のキューが選択されると前記クレジット値は1だけデクリメントされる。

【0045】さらに好ましくは、ダミーセルは実際に生成されるのではなく、各入力キューが数d_{ij}を保持する。この数は、前記クレジット値に等しく、iからjへの伝送のためにキューイングされるダミーセルの数に対応する。前記数d_{ij}及び前記クレジット値は、各時間ステップごとに、確率γ_{ij}でインクリメントされる。

【0046】さらに好ましくは、iとjの間のコネクションが許可されたときに、実セルがiとjの間の伝送のためにキューイングされている場合には、該実セルが伝送される。

【0047】もう1つの改良において、好ましくは、ダミーセルは生成されるのではなく、各入力キューが数c_{ij}を保持する。この数は、前記クレジット値に等しい。c_{ij}及び前記クレジット値は、iからjへの伝送のためにキューイングされるダミーセルの数に対応する浮動小数点数である。前記数c_{ij}及び前記クレジット値は、各時間ステップごとに、0と1の間の小数をとるγ_{ij}だけインクリメントされる。

【0048】好ましくは、iとjの間のコネクションが許可されたときに、実セルがiとjの間の伝送のためにキューイングされている場合には、該実セルが伝送される。

【0049】別の改良において、好ましくは、$c_{ij}\geqq 1$の場合、または、実セルがiとjの間のキューに存在する場合、該キューは、iとjの間でセルを伝送することができることをスケジューラに通知する。

【0050】本発明のもう1つの側面であるシューティングブランクス・スイッチは、M+N個のポートと、レート識別器と、充填決定器と、ダミーセル生成器と、スケジューラとを有する。前記ポートは、M個の入力ポート及びN個の出力ポートからなり、該M個の入力ポートと該N個の出力ポートの間でセルがルーティングされる。前記セルは、ポートのサブセットごとに異なるレートで到着する。前記M個の入力ポートはそれぞれN個の

キューを有することにより、全部でM×N個のキューが存在する。前記レート識別器は、1つのポートを、最大のセル到着レートを有するとして識別する。前記M個の入力ポートのうちの1つの入力ポートに対する到着レートは、該1つの入力ポートのN個のキューに対する到着レートの和であり、前記N個の出力ポートのうちの1つの出力ポートに対する到着レートは、前記M個の入力ポートのそれぞれから該1つの出力ポートへの到着レートの和である。前記充填決定器は、すべてのポートに対する到着レートが等しくなるように適当なキューに対するダミーフローを決定する。前記ダミーセル生成器は、前記充填決定器によって決定された到着レートでダミーセルを生成する。前記スケジューラは、キュー長を考慮しない。

【００５１】好ましくは、前記生成器は、目標到着レートが満たされるように各入力ポートにおいてダミーセルを生成し、実セルが入力ポートに到着しているときには該実セルを対応するダミーセルにピギーバックする。各キューは、該キューにおいて伝送のためにキューイングされているダミーセルの数に対応するクレジット値を保持し、スケジューラは、前記クレジット値が1以上であるときに、特定のキュー内のセルについて知らされ、該特定のキューが選択されると前記クレジット値は1だけデクリメントされる。

【００５２】好ましくは、ダミーセルは実際に生成されるのではなく、各入力キューが数d_{ij}を保持する。この数は、前記クレジット値に等しく、iからjへの伝送のためにキューイングされるダミーセルの数に対応する。前記数d_{ij}は、各時間ステップごとに、目標レートγ_{ij}に等しい確率でインクリメントされる。

【００５３】さらに好ましくは、iとjの間のコネクションが許可されたときに、実セルがiとjの間の伝送のためにキューイングされている場合には、該実セルが伝送される。

【００５４】もう1つの改良において、好ましくは、ダミーセルは生成されるのではなく、各入力キューは数c_{ij}を保持する。この数は、前記クレジット値に等しい。c_{ij}及び対応するクレジット値は、iからjへの伝送のためにキューイングされるダミーセルの数に対応する浮動小数点数である。前記数c_{ij}は、各時間ステップごとに、0と1の間の小数をとるγ_{ij}だけインクリメントされる。

【００５５】さらに好ましくは、iとjの間のコネクションが許可されたときに、実セルがiとjの間の伝送のためにキューイングされている場合には、該実セルが伝送される。

【００５６】別の改良において、好ましくは、$c_{ij} \geq 1$の場合、または、実セルがiとjの間のキューに存在する場合、該キューは、キューiとjの間でセルを伝送することができることをスケジューラに通知する。

【００５７】本発明のさらにもう1つの側面によれば、スイッチのサブセットがシューティングブランクス・スイッチであるようなスイッチを有するネットワーキングシステムが提供される。

【００５８】

【発明の実施の形態】［ＩＶＡ．シューティングブランクス・アプローチ］好ましい実施例は、最も重い負荷のポート（入力または出力のいずれでもよい）において到着レートをΛに等しくするフローの集合を提供するためのスイッチを使用する。このような場合、パケットはΛに基づく平均値で到着する。

【００５９】図3は、不平衡フローのスイッチの一例を示す模式図である。ここでは、最も重い負荷のポートはいちばん上の入力ポートである。シューティングブランクス法では、すべての入出力が同じ到着レートΛになるようにダミーパケットを生成することによって、軽い負荷のポートを通るトラフィックを増大させる。ダミーパケット生成をした後、ＭＢＭまたはＭＢＭ関連のヒューリスティックスを適用してルーティングされるべき置換を計算する。軽負荷のポートはダミートラフィックに対応する辺の端点に相当することが多いため、この方法は、マッチングに含めるようにこれらのダミーパケットを選択するようだまされることになる。

【００６０】シューティングブランクス法は、3つの主要な利点を有する。（ｉ）ＭＢＭ関連の方法を使用していながら、ＭＷＢＭの優れた安定性に匹敵することが実験的に示される（セクションＩＶＤを参照）。（ｉｉ）ＭＢＭヒューリスティックスの上にシューティングブランクスの考え方を実現することによるオーバーヘッドは小さく（セクションＩＶＢを参照）、ＭＷＢＭあるいはＭＢＭの複雑さに比べても確実に無視できる。（ｉｉｉ）このアプローチはモジュール性がある。すなわち、このアプローチは、トラフィックを「前処理」することに集中しており、任意のスケジューリング方法と組み合わせることができる。

【００６１】［ＩＶＢ．シューティングブランクスの実現］好ましい実施例の実現の詳細について以下説明する。好ましい実施例のスイッチは、サービス品質（ＱｏＳ）環境で動作し、フローは、許可された帯域を決定する契約を有する。この帯域はレートで表され、ｉ－ｊリンク、すなわち、入力ポートｉと出力ポートｊの間のリンクのレートα_{ij}は、そのリンク上のすべての顧客のために予約されたレートの和である。契約で合意されたレートは呼受付コントローラ（ＣＡＣ：Call Admission Controller）には既知である。ネットワークシステムは、どのフローも、それに割り当てられたレートを超過しないことを保証する。レートは、大きい時間ウィンドウを通じて一定であることを仮定することができるくらいまれにしか変化しない。

【００６２】本発明のシューティングブランクス法は、

3個のコンポーネントを有する。

【0063】（ｉ）充填（Fill-Up）コンポーネントでは、すべての入力ポート及び出力ポートに対するレートR_i、S_jが等しいことを保証するように、適当な入出力ペアに対してダミーフローを決定する。

【0064】（ｉｉ）生成（Generation）コンポーネントは、充填コンポーネントによって決定されたレートでダミーパケットを生成することを担当する。

【0065】（ｉｉｉ）実際のスケジューリングコンポーネントでは、選択された任意の方法（好ましくは、ＭＢＭまたはＭＢＭ関連のヒューリスティックス）を実行して、置換を計算することができる。

【0066】［ＩＶＢ．１　充填コンポーネント］図4は、図3のトラフィックに対する充填の効果の一例を示す模式図である。充填コンポーネントは、グリーディ（欲張り）手順を用いて実現される。以下、Λを入出力のレートのうちの最大値とする。このコンポーネントは新しいレートγ_{ij}を出力しようとする。γ_{ij}はオリジナルレート値α_{ij}に初期化される。グリーディ手順は入出力ポートを１からＮへの順に進む。２つのポインタが設けられ、一方は現在の入力ｉであり、他方は現在の出力ｊである。これらのポインタは、Λより小さい負荷の入力ポート及び出力ポートが見つかるまでインクリメントされる。ｉ及びｊの両方に対して、R_i，S_j＜Λである。Λ－R_i＜Λ－S_jと仮定しても一般性を失わない。レートβ_{ij}＝Λ－R_iのｉからｊへのフローを追加し、レートγ_{ij}を$\gamma_{ij}+\beta_{ij}$に更新する。入力ポインタを、$R_{i'}$＜Λであるような次のｉ′＞ｉに移動する。このステップを繰り返す。これで、充填コンポーネントの記述は完了する。

【0067】充填コンポーネントは、高々２Ｎ－１個のフローを追加することが分かる。各ステップで、一様な負荷Λを達成するという目標に向かって進むように辺を追加することが常に可能である。これは、入力レートの和が常に出力レートの和に等しいということによる。従って、この方法は正しい。

【0068】［ＩＶＢ．２　生成コンポーネント］生成コンポーネントの実現について説明する。最も重い負荷のポートのレートをΛとする。このコンポーネントは、充填コンポーネントによって決定されたレートでダミーフローを生成しなければならない。さらに複雑なことは、実フロー（例えば、入力ｉから）がその契約で合意したレート以下である場合、本発明の方法は、充填コンポーネントによってｉに対して設定されたレートΛを満たすように、ダミーパケットが生成されるレートを増大させなければならない。

【0069】生成コンポーネントに対する上記の２つの要件に効率的に対処するために、「ピギーバック」方式を用いる。入力ｉは、ちょうどΛの全レートで、充填コンポーネントによって決定された実フロー及びダミーフローの宛先へ向かうダミーパケットを生成する。すべてのスケジューラは、コネクションｉ－ｊに対して、ダミーパケットの有無に対応して１または０とみる。実パケットが入力ｉにおいて宛先をｊとしてキューイングされた場合、これは、次にスケジューラがｉ－ｊコネクションを許可するときに、ダミーパケットの先頭に移される。ピギーバック方式によれば、シューティングブランクスの実現は、実フローの高価なポリシング(policing)をさらに行うことなしに、しかも、任意のリンクにおいてダミーパケット生成に固定レートを使用しながら（これはハードウェアで容易に実現可能な方法である。）達成される。

【0070】さらに効率を改善するため、さらに好ましい実施例では、ダミーパケットは、明示的に生成され保持されるのではない。入出力ポートのあらゆるペア（ｉ，ｊ）に対して、ｉからｊへの伝送のためにキューイングされているダミーパケットの数d_{ij}のみが保持される。この数は、各時間ステップごとに、確率γ_{ij}でインクリメントされる。d_{ij}＞０の場合、スケジューラは１（すなわち、ｉ－ｊ辺が２部グラフ内に存在する）と判断し、そうでない場合は０と判断する。スケジューラがｉ－ｊコネクション（これはd_{ij}＞０を意味する。）を許可し、q_{ij}＞０（これは、伝送のためにキューイングされている実パケットがあることを意味する。）である場合、実パケットがルーティングされ、d_{ij}は１だけデクリメントされる。ｉ－ｊコネクションが許可され、実パケットがキューイングされていない場合、単に、d_{ij}が１だけデクリメントされる。

【0071】図5は、本発明によるさらに好ましい実施例であり、生成コンポーネントのカウンタベースでの実現を示す模式図である。この実施例では、確率的アプローチではなく決定性アプローチを使用する。このアプローチは、非常にコストのかかる乱数発生器の使用を回避する。図5において、ｉ－ｊコネクションに対して充填コンポーネントにより決定されたレートをγ_{ij}とする。カウンタc_{ij}が設けられ、各時間ステップごとにγ_{ij}（０と１の間の小数をとる。）だけインクリメントされる。カウンタの値は、負の値もとりうる浮動小数点数である。前と同様に、整数q_{ij}は実パケットの数を保持する。スケジューラは、q_{ij}＞０またはc_{ij}≧１の場合、２部グラフ内にｉ－ｊ辺があると判断する。スケジューラがｉ－ｊコネクションを許可する場合、c_{ij}は１だけデクリメントされ、実パケットがあれば（すなわちq_{ij}＞０）実パケットがルーティングされる。

【0072】［ＩＶＣ．１　シューティングブランクス・スイッチ］図6は、本発明によるシューティングブランクス・スイッチの好ましい実施例を示す概略的構成図である。シューティングブランクス・スイッチは、Ｍ個の入力ポート６．１１～６．１４及びＮ個の出力ポート６．２１～６．２４を有するものとする。即ち、図示し

たスイッチではM＝４及びN＝４である。セルは、入力ポートに到着した後、M個の入力ポートとN個の出力ポートの間でルーティングされなければならない。これらのセルは、相異なるレートで到着する。各セルは、いくつかのキュー（６．７１１～６．７１４は一部のキューである。）のうちの１つにキューイングされる。１つのキューは、特定の入力ポートから特定の出力ポートへルーティングされるセルに対応する。例えば、キュー６．７１４は、入力ポート６．１１から出力ポート６．２４へルーティングされるセルをキューイングする。同様に、キュー６．７１２は、入力ポート６．１１から出力ポート６．２２へルーティングされるセルをキューイングする。従って、各入力ポートは、その入力ポートからN個の出力ポートのそれぞれへルーティングされるセルに対応するN個のキューを有する。こうして、このスイッチにはM×N個のキューがある。この図の場合では４×４＝１６個のキューがある。

【００７３】レート識別器６．３は、最大到着レートのポートを識別する。入力ポートに対する到着レートは、その入力ポートのN個のすべてのキューに対する到着レートの和である。同様に、出力ポートに対する到着レートは、M個の入力ポートのそれぞれからその出力ポートへの到着レートの和である。充填決定器６．４は、すべてのポートに対する到着レートが等しくなるように適当なキューに対するダミーフローを決定する。ダミーセル生成器６．５は、充填決定器によって決定された到着レートでダミーセルを生成する。スケジューラ６．６は、キュー長を考慮しないヒューリスティックMBM法に基づいてルーティングを行う。

【００７４】［ＩＶＣ．２　シューティングブランクス・スイッチを用いたネットワークシステム］図7は、シューティングブランクス・スイッチを用いたネットワークシステムの好ましい例を示す。ネットワークシステム７．１は、いくつかのスイッチを有する。これらのスイッチのサブセットが、シューティングブランクス・スイッチ７．２１～７．２３からなる。シューティングブランクス・スイッチの好ましい実施例の構造は前のセクションＩＶＣ．１（図６）に記載されている。

【００７５】［ＩＶＤ．実験結果］注意すべき点であるが、ここに記載する実験は単なる例示であり、本発明の技術的範囲を制限するものと解釈してはならない。これらのシミュレーション実験を用いて、シューティングブランクス法が、シューティングブランクス法を使用しない従来のスケジューリングアプローチに比べて優れた結果を与えることを示す。強調されるべき点であるが、シューティングブランクス法は、任意のスケジューリングヒューリスティックスと直交して組み合わせることが可能であるが、平衡入力に最大サイズマッチングに近似する良好な方法と組み合わせるほうが良好に動作する。ここに記載する実験は、以下のようにして生成されるトラフィックパターン例を使用する。任意の時間ステップにおいて、各入力ポートｉは確率$\Sigma_{j=1}^{N}\alpha_{ij}$で１個のセルを生成する。入力ポートｉで生成されたセルの宛先が出力ポートｊである確率は$\alpha_{ij}/\Sigma_{j=1}^{N}\alpha_{ij}$である。

【００７６】入力ポートの次数とは、それがマッチングされる出力ポートの数のことである。同様に、出力ポートの次数とは、それがマッチングされる入力ポートの数である。前述の米国特許出願及び特願平１１－２４４６４８号に記載されているような低次数で不平衡のトラフィックパターンは、例とするトラフィックタイプとみなされる。このトラフィックタイプは、McKeown et al.に記載されたトラフィックパターンの一般化である。McKeownに記載されたトラフィックパターンは許容されるが、スケジューリング方法がＭＢＭである場合にはキューが無制限に成長する（Nick McKeown, Venkat Anantharam, and Jean Walrand, "Achieving 100%throughput in an input-queued switch", In Proceedings IEEE INFOCOM' 96,p.296-302, San Francisco, CA, March 1996、を参照）。今のシミュレーションの場合、半数の入力ポート及び半数の出力ポートがちょうど１個のフローをサポートし、他の入力及び出力ポートはちょうど２個のフローをサポートする。各フローは同じレートでトラフィックを生成するため、半数のポートはレートλで負荷を受け、他の半数は２λで負荷を受ける。シミュレーションの開始時のフローを生成するため、２個のランダム置換を使用して、Ｎ！個の可能性のそれぞれを等確率で選択する。重い負荷の入力ポートに対しては、置換によって定義される両方のフローを使用する。軽い負荷の入力ポートに対しては、フローを定義するのに第１の置換のみを用いる。

【００７７】サイズN＝４及びN＝８のスイッチについて考える。完全を期すため、使用する実際のレート行列を示す。（ｉ，ｊ）成分は、入力ポートｉから出力ポートｊへのレートを表す。実トラフィックに対するレート行列をα_Nで表す。ブランク（ダミー）に対するレート行列をβ_Nで表す。全レート行列を$\gamma_N=\alpha_N+\beta_N$で表す。なお、$\alpha_N$は、前段落に記載したように生成され、$\beta_N$は、前述のセクションＩＶＢ．１に記載したように$\alpha_N$から計算される。

【００７８】サイズN＝４の場合、$\gamma_4=\alpha_4+\beta_4$、すなわち、

【００７９】

【数１】

$$\begin{bmatrix} 0 & \lambda & \lambda & 0 \\ 0 & \lambda & 0 & \lambda \\ \lambda & 0 & \lambda & 0 \\ \lambda & 0 & 0 & \lambda \end{bmatrix} = \begin{bmatrix} 0 & 0 & \lambda & 0 \\ 0 & \lambda & 0 & 0 \\ \lambda & 0 & \lambda & 0 \\ \lambda & 0 & 0 & \lambda \end{bmatrix} + \begin{bmatrix} 0 & \lambda & 0 & 0 \\ 0 & 0 & 0 & \lambda \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

である。

【0080】サイズN＝8の場合、$\gamma_8=\alpha_8+\beta_8$、すなわち、

【0081】

【数2】

$$\begin{bmatrix} 0 & \lambda & 0 & 0 & 0 & 0 & \lambda & 0 \\ \lambda & 0 & \lambda & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda & \lambda & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda & \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda & 0 & \lambda & 0 \\ 0 & \lambda & 0 & 0 & 0 & 0 & 0 & \lambda \\ \lambda & 0 & 0 & 0 & 0 & \lambda & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda & 0 & \lambda \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \lambda & 0 \\ \lambda & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda & 0 & \lambda & 0 \\ 0 & \lambda & 0 & 0 & 0 & 0 & 0 & \lambda \\ \lambda & 0 & 0 & 0 & 0 & \lambda & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda & 0 & \lambda \end{bmatrix} + \begin{bmatrix} 0 & \lambda & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

である。

【0082】$0 \leqq \lambda \leqq 1/2$の範囲を考える。任意のポートに対する最大負荷は$2\lambda$であるため、$2\lambda$をシミュレーションのレートと呼ぶ。

【0083】実験は、特定のコンフィグレーションに対する安定性の最大レートを経験的に評価しようとするものである。ただし、コンフィグレーションは、スイッチサイズ、スケジューリング方法、及びトラフィックパターンからなる。正式な用語ではないが、入力キューが無制限に成長しない場合に、コンフィグレーションは安定であるという。安定性の経験的定義は以下の通りである：入力キューサイズの総和が、ある一定の時間ステップ数Tのうちに、ある一定の上限Qより大きく成長しない場合、コンフィグレーションは安定であると宣言される。なお、Qの増大あるいはTの減少は、一部のコンフィグレーションの実験的安定性の範囲を増大させる効果を有することがある。また、QまたはTのいずれかを増大させることは、シミュレーション時間を増大させることがある。我々の実験では、$Q=16N^2$個のセルと、$T=256N^2$を使用する。安定性実験は、我々の安定性基準が破られるまでにコンフィグレーションが動作可能な最大レートを決定する。このような安定性実験は非常に長くなることがあるため、離脱条件を追加する。1，000ステップごとにキューサイズの総和をチェックする。この総和が前回のチェック以来成長していない場合、システムは安定であると仮定される。

【0084】表1に実験結果を示す。9個のスケジューリング方法に対して、安定性のレートが、ブランクなしの4個のポート、ブランクありの4個のポート、ブランクなしの8個のポート、ブランクありの8個のポートに対してテストされている。スケジューリング方法は、3つのクラス、すなわち、MBM、高品質MBMヒューリスティックス、及び低品質MBMヒューリスティックスにグループ分けされている。ヒューリスティックスの品質とは、ブランクを追加する前のもとのトラフィックに対するMBMの挙動をどのくらいよく近似するかをいう。安定性レートがΛより低いヒューリスティックスは、ここでの考察の目的では低品質として特徴づけられる。反復法はすべて4回（4イタレーション）実行した。

【００８５】主要な結果は表の第１行に示されている。Ｎ＝４の場合、ＭＢＭは、ブランクなしでΛまで安定であり、ブランクありで１００％まで安定である。Ｎ＝８の場合、ブランクの使用は安定性レートを９３％から１００％に上昇させる。

【００８６】次のスケジューリング方法のクラス（これは、ＰＩＭ－ＳＨＡＫＥ－４、ＲＮＤ－ＳＨＡＫＥ－４、ＲＲＲＮＤ－ＳＨＡＫＥ－４、及びＳＬＩＰ－ＳＨＡＫＥ－４を含む。）は、高品質ヒューリスティックスとみなされる。このクラスのヒューリスティックスの場合、ブランクの使用により全般的に改善が見られる。代表的な改善は、４ポートでは安定性９３％から安定性９８％への改善であり、８ポートでは安定性９０％から安定性９２％への改善である。なお、これらの方法が、Ｎ＝４のブランクなしの場合にＭＢＭよりも良好に動作することは、この例については説明することができるが、一般には成り立たない。これらのヒューリスティックスは、２部グラフにおいて接続している辺が多いノードのほうに偏る。これは、これらの例にとって有利である。というのは、２本の辺を有するノードは２倍の帯域をサポートする必要があるからである。

【００８７】

【表１】

スケジューリングアルゴリズム	４ポート		８ポート	
	安定レート	安定レート ブランク有	安定レート	安定レート ブランク有
MBM	90%	100%	93%	10%
PIM-SHAKE-4	93%	98%	90%	92%
RND-SHAKE-4	92%	98%	90%	92%
RR-RND-SHAKE-4	93%	98%	89%	92%
SLIP-SHAKE-4	93%	100%	93%	100%
PIM-4	88%	86%	87%	86%
RND-4	88%	86%	87%	86%
RR-RND	88%	86%	88%	85%
SLIP-4	100%	80%	100%	86%

【００８８】最後のスケジューリング方法のクラスは、比較的低品質のＭＢＭヒューリスティックスである。このクラスは、ＰＩＭ－４、ＲＮＤ－４、ＲＲ－ＲＮＤ、及びＳＬＩＰ－４を含む。一般に、これらの方法は、ブランクを使用するとわずかにパフォーマンスが悪くなり、いずれのスイッチサイズの場合でも安定性レートが８８％から８６％に低下する。ＳＬＩＰ－４は例外である。ＳＬＩＰ－４の場合の１００％のスループットは、我々のトラフィックフロー例が与えられれば、予測することができる。優先度ホイールが完全に同期することにより、１００％のスループットが可能となる。しかし、ＳＬＩＰ－４は、この例における平衡トラフィックでは問題がある。一般に、ＳＬＩＰは、多くの低次数トラフィックパターンに対して高品質の解を達成するのは困難であることを示すことができる（前述の米国特許出願及び特願平１１－２４４６４８号を参照）。この実験においてブランクなしのＳＬＩＰの良好なパフォーマンスは、単純なトラフィックパターンによる見かけ上の結果である。

【００８９】本発明の他の変形例は、当業者には以上の説明から明らかである。本発明のいくつかの実施例についてのみ具体的に説明したが、本発明の技術思想及び技術的範囲から離れることなく、さまざまな変形が可能であることは明らかである。

【００９０】

【発明の効果】以上詳細に説明したように、本発明によるスケジューリング方法によれば、まず、１つのポートを最大のセル到着レートを有するとして識別し、次に全てのポートに対する到着レートがほぼ等しくなるように、最大のセル到着レートを有するとして識別されたポート以外のすべてのポートにダミーセルを追加する。こうすることで、ＭＷＢＭ（最大重み２部マッチング）スケジューラの優れた安定性を達成することができるとともに、より単純なＭＢＭ（最大サイズ２部マッチング）またはＭＢＭベースのヒューリスティックスを実行することができる。

【図面の簡単な説明】

【図１】スケジューリング問題の２部グラフを説明するための模式図である。

【図２】仮想出力キューイング（ＶＯＱ）をサポートするＮ＝４のクロスバスイッチの模式的構成図である。

【図３】不平衡フローのスイッチの一例を示す模式図である。

【図４】本発明の一実施形態における充填コンポーネントによるダミーフローの生成後の、図３のトラフィックの新しいレートを示す模式図である。

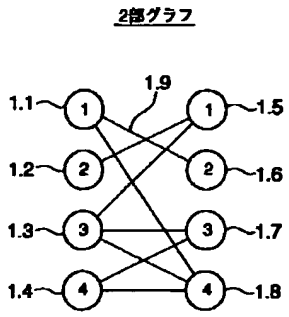
【図５】本発明によるさらに好ましい実施形態であり、生成コンポーネントのカウンタベースでの実現を示す模式図である。

【図６】本発明によるシューティングブランクス・スイッチの好ましい実施例を示す概略的構成図である。

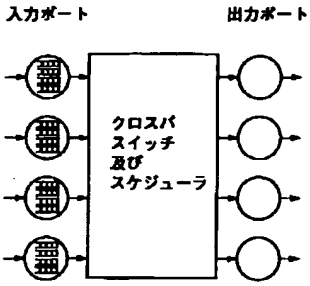
【図７】本発明によるシューティングブランクス・スイッチを用いたネットワークシステムの好ましい実施例を示す模式図である。

【符号の説明】
6．11～6．14　入力ポート
6．21～6．24　出力ポート
6．3　レート識別器
6．4　充填決定器
6．5　ダミーセル生成器
6．6　スケジューラ
6．711～6．714　キュー
7．1　ネットワークシステム
7．21～7．23　シューティングブランクススイッチ

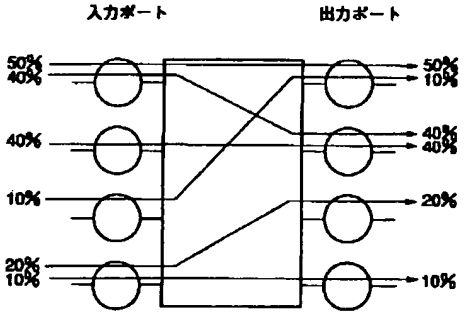
【図1】



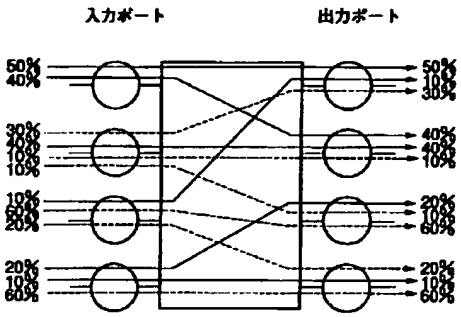
【図2】



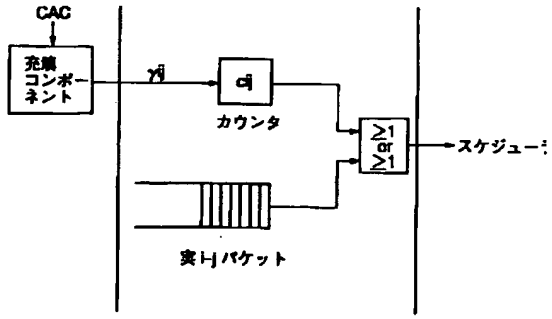
【図3】



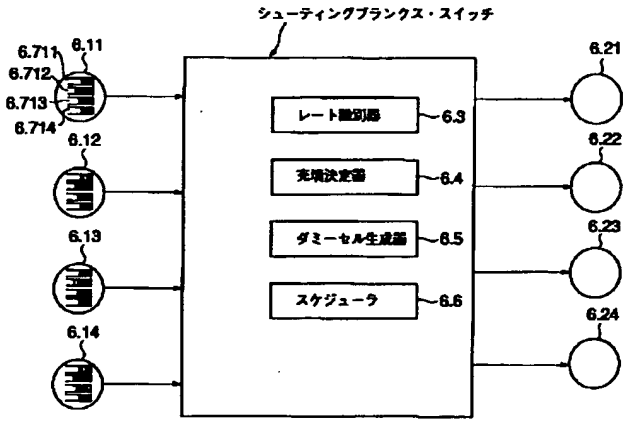
【図4】



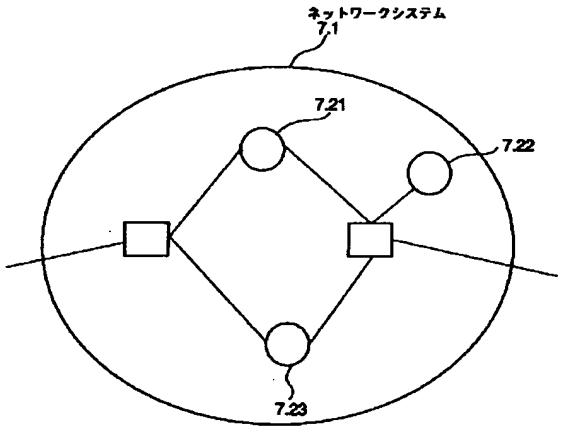
【図5】



【図６】



【図７】



【手続補正書】
【提出日】平成１１年９月１日（１９９９．９．１）
【手続補正１】
【補正対象書類名】明細書
【補正対象項目名】発明の名称
【補正方法】変更
【補正内容】
【発明の名称】　キューイングシステムにおけるスケジューリング方法及びシューティングブランクス・スイッチ

フロントページの続き

(72)発明者　スタフロス・コリオパロス
アメリカ合衆国、ニュージャージー　08540　プリンストン、４　インディペンデンス　ウエイ、エヌ・イー・シー・リサーチ・インスティテュートゥ・インク内

(72)発明者　サティシ・ラオ
アメリカ合衆国、ニュージャージー　08540　プリンストン、４　インディペンデンス　ウエイ、エヌ・イー・シー・リサーチ・インスティテュートゥ・インク内

Fターム(参考) 5K030 GA03 HA10 HB09 HB17 KA03
KA13 KX12 KX18 KX29 LA03
LB07 LE05 MA13
9A001 CC03 CC07 FZ05 HH34 KK56
LL09